

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平 10 - 97285

(43) 公開日 平成 10 年 (1998) 4 月 14 日

(51) Int. Cl. °	識別記号	庁内整理番号	F I	技術表示箇所
G10L 3/00	561		G10L 3/00	561 G
	531			531 E

審査請求 未請求 請求項の数 9 O L (全 14 頁)

(21) 出願番号 特願平 8 - 251373

(22) 出願日 平成 8 年 (1996) 9 月 24 日

(71) 出願人 000006013
三菱電機株式会社
東京都千代田区丸の内二丁目 2 番 3 号

(72) 発明者 岩▲崎▼ 知弘
東京都千代田区丸の内二丁目 2 番 3 号 三
菱電機株式会社内

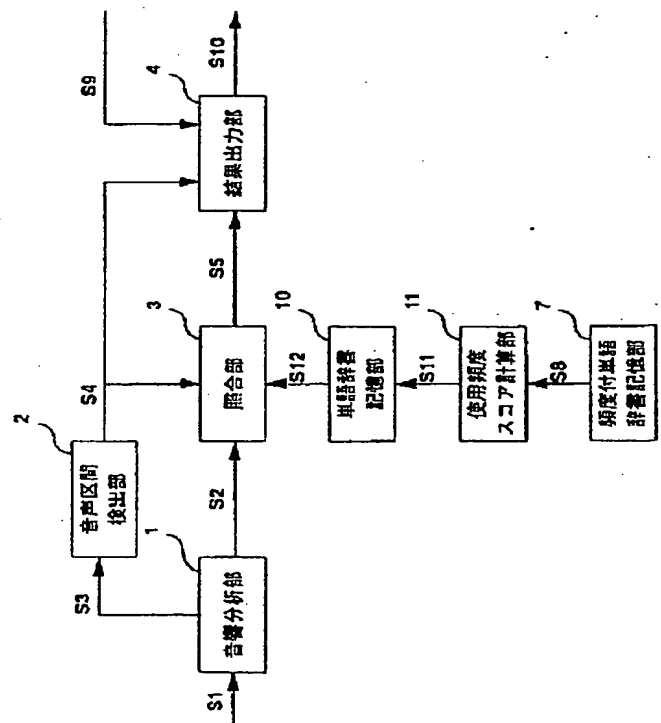
(74) 代理人 弁理士 宮田 金雄 (外 3 名)

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】 音声認識装置において、大語彙の音声が入力される場合でも高い精度で音声を認識する。

【解決手段】 頻度付単語辞書記憶部 7 の使用頻度を表す情報から使用頻度の高いものほど小さな値となる使用頻度スコアを計算し、単語辞書記憶部 10 に記憶する使用頻度スコア計算部 11 を備え、照合部 3 において入力された音声信号 S1 と単語辞書 S12 が音響的にどの程度近いかを示す音響スコアに単語辞書記憶部に 10 に記憶されているその単語の使用頻度スコアを規定の割合で加算して距離値 S5 とする。かくして、使用頻度が高い単語ほど小さな値となる使用頻度スコアを与え、パターン照合時にこれを加えることにより、使用頻度の高い単語の認識率できる。



【特許請求の範囲】

【請求項 1】 入力される音声信号を一定時間毎に音響分析し、特徴パラメータベクトルと音声信号のパワーとに順次変換し出力する音響分析部と、当該音響分析部から受け取る上記音声信号のパワーの変化により上記音声信号の音声区間を検出し、当該音声区間の検出状況の変化により音声区間検出信号を出力する音声区間検出部と、上記音声区間検出信号の指示に従い上記音響分析部から出力される上記特徴パラメータベクトルの時系列のうち上記音声区間の部分と単語辞書記憶部に記憶されている単語辞書との照合を行い、入力された上記音声信号と上記単語辞書とのパターン照合を行い、距離値として出力する照合部と、上記音声区間検出信号により指示された時、既に受け取った上記距離値をソーティングして上記距離値の小さな 1 つ又は複数の単語を認識結果として出力する結果出力部とを有する音声認識装置において、認識対象単語の標準パターンと単語の使用頻度を表す情報を収める頻度付単語辞書記憶部と、当該頻度付単語辞書記憶部の使用頻度を表す情報から使用頻度の高いものほど小さな値となる使用頻度スコアを計算し、上記単語辞書記憶部に記憶する使用頻度スコア計算部とを備え、上記照合部において入力された上記音声信号と単語辞書が音響的にどの程度近いかを示す音響スコアに上記単語辞書記憶部に記憶されているその単語の上記使用頻度スコアを規定の割合で加算して距離値とすることを特徴とする音声認識装置。

【請求項 2】 上記使用頻度スコア計算部は上記使用頻度スコアが規定の下限値より小さくならないように設定することを特徴とする請求項 1 に記載の音声認識装置。

【請求項 3】 既存のデータベースから同じふり仮名を持つ単語の頻度を計数し、全体に対する上記単語の頻度を使用頻度とみなす使用頻度推定部を備えることを特徴とする請求項 1 に記載の音声認識装置。

【請求項 4】 上記使用頻度推定部はふり仮名をローマ字表記したものが「OU」を含む時、規定の割合で「OO」に置き換えた読みがされるとし、元の上記単語の使用頻度を上記規定の割合で減ずると共に、新たに上記「OU」を上記「OO」で置き換えた単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の上記規定の割合とすることを特徴とする請求項 3 に記載の音声認識装置。

【請求項 5】 上記使用頻度推定部はふり仮名をローマ字表記したものが「EI」を含む時、規定の割合で「EE」に置き換えた読みがされるとし、元の上記単語の使用頻度を上記規定の割合で減ずると共に、新たに上記「EI」を上記「EE」で置き換えた単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の上記規定の割合とすることを特徴とする請求項 3 に記載の音声認識装置。

【請求項 6】 上記使用頻度推定部は任意の規定の割合

で母音及び撥音が長音化されるものとし、元の上記単語の使用頻度を上記規定の割合で減ずると共に、新たに上記母音及び撥音を長音化したもので置き換えた単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の上記規定の割合とすることを特徴とする請求項 3 に記載の音声認識装置。

【請求項 7】 上記使用頻度推定部は任意の規定の割合で音節毎に区切られるものとし、元の単語の使用頻度を上記規定の割合で減ずると共に、上記音節毎に区切られた単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の上記規定の割合とすることを特徴とする請求項 3 に記載の音声認識装置。

【請求項 8】 上記使用頻度推定部は任意の規定の割合で促音が「つ」と発声されるものとし、元の単語の使用頻度を上記規定の割合で減ずると共に、促音が「つ」と発声された単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の上記規定の割合とすることを特徴とする請求項 3 に記載の音声認識装置。

【請求項 9】 上記使用頻度推定部は規定の基準においてデータベースの内容を分類して、それぞれの分類毎に使用頻度を推定し、上記使用頻度スコア計算部は上記分類毎に使用頻度スコアを演算し、また同じ規定の基準において分類された話者の音声信号より学習された標準パターンをもって、未知の話者の音声信号の話者識別を行い上記話者がどの分類に近いかを示す話者識別スコアを出力する話者識別部を備え、上記照合部は上記話者識別スコアと当該分類における単語の使用頻度スコアと単語の音響スコアを任意の規定の割合で加算して照合結果とすることを特徴とする請求項 3 に記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は音声認識装置に関し、大語彙を単語を対象して認識するものに適用し得る。

【0002】

【従来の技術】住所や姓名のような大語彙を対象とする音声認識装置では、類似する単語が多くなるため認識性能が低下し、また大語彙とのパターン照合を行うため演算量が膨大となる問題があり、実現が極めて難しいものである。従来、大語彙を対象とするこの種の音声認識装置として、特開平 3 - 8 4 6 0 0 号公報に開示されたものを図 9 に示す。

【0003】音響分析部 1 は入力される音声信号 S 1 を一定時間毎に音響分析し、特徴パラメータベクトル S 2 と音声信号のパワー S 3 に変換し出力する。音声区間検出部 2 は音響分析部 1 から受け取る音声信号のパワー S 3 の変化により音声信号の音声区間を検出し、音声区間の検出状況の変化により音声区間検出信号 S 4 を出力する。照合部 3 は音声区間検出信号 S 4 の指示に従い、音

響分析部 1 から受け取る特徴パラメータベクトル S 2 の時系列のうち音声区間のものと単語辞書記憶部 5 から読み出される順番で単語辞書 S 5 との照合を行い、入力された音声信号 S 1 と単語辞書 S 6 が音響的にどの程度近いを示す音響スコアを距離値 S 5 として順次出力する。

【0004】なお頻度付単語辞書記憶部 7 は認識対象単語の読みを表すラベルと使用頻度を表す情報を収め、単語辞書ソーティング部 6 は頻度付単語辞書記憶部 7 の使用頻度の高い順番に単語情報を並び替え、単語辞書記憶部 5 は並び替えた単語情報を記憶する。また図中、S 6 は単語辞書、S 7 は単語辞書、S 8 は頻度付き単語辞書である。結果出力部 4 は音声区間検出信号 S 4 又は外部から入力される出力要求信号 S 9 が入力された時、既に受け取った距離値 S 5 のうちまだ出力していないものを距離値 S 5 によりソーティングして距離値 S 5 の小さな 1 つ又は複数の単語を認識結果 S 10 として出力する。

【0005】このような構成の音声認識装置の動作について説明する。認識に先だって単語辞書ソーティング部 6 では頻度付単語辞書記憶部 7 の内容を読み出し、使用頻度によりソーティングを行い、使用頻度の高い順番に単語辞書記憶部 5 に収める。以下、認識時の動作について説明する。認識装置は 10m 秒程度の時間を単位として処理が進められる。この単位時間をフレームと呼ぶ。音響分析部 1 はフレーム毎に入力された音声信号 S 1 を音響分析し、特徴パラメータベクトル S 2 と音声信号のパワー S 3 に変換する動作を繰り返す。音響分析の手法としては、例えば LPC (Linear Prediction coefficient) 分析や FFT (高速フーリエ変換)、フィルタバンクによる手法等が用いられる。

【0006】次に音声区間検出部 2 の動作を説明する。音声区間検出部 2 では音声区間の検出は音声信号のパワー S 3 を監視し、音声信号のパワー S 3 がある閾値を越えたら音声区間の始端とし、閾値より下回ったら音声区間の終端候補とし、そのまま閾値以下で一定時間継続すると終端候補が正しかったものとして終端確定する。この時間は一般的には 0.3 秒程度が適当とされている。0.3 秒以内に再び閾値を越えて立ち上がると、先ほど検出した終端候補を無効とする。

【0007】具体的に例をあげて音声区間検出部 2 の動作を、図 10 を用いて説明する。図 10 において「ほった」という発声の音声信号のパワーの変化の一例を示す。横軸は時間、縦軸は音声信号のパワーの大きさを表す。フレーム T 1 からフレーム T 2 までが「ほ」、フレーム T 2 からフレーム T 3 までが「っ」、フレーム T 3 からフレーム T 4 までが「た」の発声区間を想定している。図 10 では音声信号のパワーは雑音レベルからフレーム T 1 時点で閾値 P 1 を越えフレーム T 2 で下回る。再びフレーム T 3 で閾値 P 1 を越えフレーム T 4 で下回る。フレーム T 5 はフレーム T 4 から 0.3 秒経過した時

点を指す。「っ」は音響的に促音に分類される。普通の発声では促音は 0.3 秒以下の時間長となるため、この例でもフレーム T 2 とフレーム T 3 の間は 0.3 秒以下の時間とする。上述の音声区間検出部 2 の動作に従えば、フレーム T 1 からフレーム T 4 を音声区間として検出する。

【0008】音声区間検出部 2 では音声区間検出信号 S 4 として始端信号、終端候補信号、終端確定信号の 3 種類を送出する。図 10 ではフレーム T 1 とフレーム T 3 で始端信号を、フレーム T 2 とフレーム T 4 で終端候補信号を、フレーム T 5 で終端確定信号を送出する。終端候補信号の後、終端確定信号が送出されずに、始端信号が送出された場合は、その前の終端候補信号の終端候補、つまりフレーム T 2 を無効とすることを示す。

【0009】照合部 3 では音響分析部 1 から音声区間の特徴パラメータベクトル S 2 が送られてくるので、音声区間検出信号 S 4 で指定される始端信号から終端確定信号までの間、内部に蓄える。音声区間検出部 2 から音声区間検出信号 S 4 として終端候補信号を受信したならばパターン照合を開始する。図 10 にパターン照合を行うフレームを斜線で示す。パターン照合の方法はさまざまあるが、例えば DP (Dynamic Programming) マッチングや HMM (Hidden Markov Model) による方法が適用できる。照合部 3 は単語辞書記憶部 5 の単語辞書 S 6 を並べられている順番に読み出し、内部に蓄えられているフレーム T 1 からフレーム T 2 の区間の特徴パラメータベクトル S 2 に対するパターン照合を行い、距離値 S 5 を結果出力部 4 に送出する。単語辞書記憶部 5 のの中では単語辞書 S 6 は頻度の高い順番に並べられているため、パターン照合は頻度の高い単語から行われることになる。

【0010】続いてフレーム T 3 で音声区間検出部 2 から音声区間検出信号 S 4 として始端信号を受信した時には、フレーム T 1 からフレーム T 2 までの区間が無効であるとしパターン照合を中止する。さらに続いてフレーム T 4 で音声区間検出部 2 から音声区間検出信号 S 4 として終端候補信号を受信した時、内部に蓄えているフレーム T 1 からフレーム T 4 の区間の特徴パラメータベクトル S 2 に対するパターン照合を行い、入力された音声信号 S 1 と単語辞書 S 6 がどの程度近いを示す音響スコアとその単語を距離値 S 5 として順次、結果出力部 4 に出力する。

【0011】結果出力部 4 では照合部 3 から送られてくる距離値 S 5 に対しその音響スコアにより順次ソーティングを行う。音声区間検出部 2 からの音声区間検出信号 S 4 として始端信号を受けた時、それまでにソーティングされた距離値 S 5 をクリアする。音声区間検出部 2 からの音声区間検出信号 S 4 として終端確定信号を受けた時、それまでにソーティングされた距離値 S 5 のうち上位から 1 つ又は複数の認識結果 S 10 を出力する。この

出力結果を画面に出力したものを発声者が確認し、正しい認識結果が含まれていない場合には出力要求信号 S 9 を入力する。この出力要求信号 S 9 が入力された場合には、その時点までにソーティングされた距離値 S 5 のうちで、まだ出力していないもののうち上位から 1 つ又は複数の単語を認識結果 S 10 として出力する。このなかに認識結果が含まれていない場合には、さらに上記のシーケンスが繰り返される。

【0012】結果出力部 4 の処理の流れを図 10 を用いてさらに説明する。フレーム T 1 で結果出力部 4 は内部のデータをクリアする。フレーム T 2 から照合部 3 から距離値 S 5 が転送されてくるので順次その距離値 S 5 によりソーティングを行い内部に蓄える。フレーム T 3 でソーティングした結果をクリアする。フレーム T 4 で再び照合部 3 から距離値 S 4 が転送されてくるので順次その音響スコアによりソーティングを行い内部に蓄える。フレーム T 5 でソーティングされた距離値 S 5 のうち上位から 1 つ又は複数の単語を認識結果 S 10 として出力する。図 10 に認識結果 S 10 を出力している期間を黒く塗った長方形で示している。フレーム T 5 以降も照合部 3 から距離値 S 5 が転送されてくるので順次ソーティングを行い内部に蓄える。フレーム T 6 で外部から出力要求信号 S 9 が入力されるので、その時点までソーティングされた距離値 S 5 のうち上位から 1 つあるいは複数の単語を認識結果 S 10 として出力する。

【0013】上記のように、従来の技術による音声認識装置では、大語彙の単語認識を行う場合に頻度の高い単語の認識結果は発声終了後 0.3 秒で出力される。また、頻度の低い単語もしばらく後に装置に出力要求信号 S 9 を送ることで認識結果を得ることができる。

【0014】

【発明が解決しようとする課題】ところが従来の音声認識装置は以上のように構成されているので、どのように丁寧に発声しても頻度の低い単語は発声終了後 0.3 秒以内では認識できない。また、姓名のように数万単語という大語彙を認識しようとする場合、「大野／小野」、「佐藤／里」など類似した単語が増加するため、認識率が低下するという問題があった。図 11 に電話回線を通して収集した姓を発声する音声データに対する従来の音声認識装置の認識性能を示す。図中縦軸は誤り率、横軸はパターン照合に用いた頻度の高い単語数を対数 (log) で示す。図中実線で示したものが従来の音声認識装置による誤り率であり、点線はパターン照合に用いた単語辞書記憶部 5 の単語のなかに正解が含まれていなかった割合を示す。

【0015】日本人の姓の総数は約 58,000 単語であり、右に行くほど多くの単語とパターン照合を行っている。頻度の高い 1,000 単語を対象に認識する場合は、正解がこの 1,000 単語の中に含まれない割合である脱落率が 30.7 店=誤認識の 16.5% を合わせて 47.2% の発声が不正解

となるが、さらに時間をかけて照合を行い 57,711 単語を照合させた場合、1.3% の脱落と 62.1% の誤認識を合わせて、63.4% が不正解となることを示している。つまり、単語数を増やすと極端に認識率が低下し、結果として不正解が増加する問題があった。

【0016】この発明は上記のような問題点を解消するためになされたもので、大語彙の場合でも高い精度で音声認識し得る音声認識装置を提供するものである。

【0017】

10 【課題を解決するための手段】この発明に係る音声認識装置は、入力される音声信号を一定時間毎に音響分析し、特徴パラメータベクトルと音声信号のパワーとに順次変換し出力する音響分析部と、その音響分析部から受け取る音声信号のパワーの変化により音声信号の音声区間を検出し、その音声区間の検出状況の変化により音声区間検出信号を出力する音声区間検出部と、音声区間検出信号の指示に従い音響分析部から出力される特徴パラメータベクトルの時系列のうち音声区間の部分と単語辞書記憶部に記憶されている単語辞書との照合を行い、
20 入力された音声信号と単語辞書とのパターン照合を行い、距離値として出力する照合部と、音声区間検出信号により指示された時、既に受け取った距離値をソーティングして距離値の小さな 1 つ又は複数の単語を認識結果として出力する結果出力部とを有する音声認識装置において、認識対象単語の標準パターンと単語の使用頻度を表す情報を収める頻度付単語辞書記憶部と、その頻度付単語辞書記憶部の使用頻度を表す情報から使用頻度の高いものほど小さな値となる使用頻度スコアを計算し、単語辞書記憶部に記憶する使用頻度スコア計算部とを備え、
30 照合部において入力された音声信号と単語辞書が音響的にどの程度近いかを示す音響スコアに単語辞書記憶部に記憶されているその単語の使用頻度スコアを規定の割合で加算して距離値とするものである。

【0018】また次の発明に係る音声認識装置は、使用頻度スコア計算部は使用頻度スコアが規定の下限値より小さくならないように設定するものである。

【0019】また次の発明に係る音声認識装置は、既存のデータベースから同じふり仮名を持つ単語の頻度を計数し、全体に対する単語の頻度を使用頻度とみなす使用頻度推定部を備えるものである。

【0020】また次の発明に係る音声認識装置は、使用頻度推定部はふり仮名をローマ字表記したものが「OU」を含む時、規定の割合で「OO」に置き換えた読みがされるとし、元の単語の使用頻度を規定の割合で減ずると共に、新たに「OU」を「OO」で置き換えた単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【0021】また次の発明に係る音声認識装置は、使用頻度推定部はふり仮名をローマ字表記したものが「EI」を含む時、規定の割合で「EE」に置き換えた読み

がされるとし、元の単語の使用頻度を規定の割合で減ずると共に、新たに「E I」を「E E」で置き換えた単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【0022】また次の発明に係る音声認識装置は、使用頻度推定部は任意の規定の割合で母音及び撥音が長音化されるものとし、元の単語の使用頻度を規定の割合で減ずると共に、新たに母音及び撥音を長音化したもので置き換えた単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【0023】また次の発明に係る音声認識装置は、使用頻度推定部は任意の規定の割合で音節毎に区切られるものとし、元の単語の使用頻度を規定の割合で減ずると共に、音節毎に区切られた単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【0024】また次の発明に係る音声認識装置は、使用頻度推定部は任意の規定の割合で促音が「つ」と発声されるものとし、元の単語の使用頻度を規定の割合で減ずると共に、促音が「つ」と発声された単語を加え、その

【0025】また次の発明に係る音声認識装置は、使用頻度推定部は規定の基準においてデータベースの内容を分類して、それぞれの分類毎に使用頻度を推定し、使用頻度スコア計算部は分類毎に使用頻度スコアを演算し、また同じ規定の基準において分類された話者の音声信号より学習された標準パターンをもって、未知の話者の音声信号の話者識別を行い話者がどの分類に近いかを示す話者識別スコアを出力する話者識別部を備え、照合部は話者識別スコアとその分類における単語の使用頻度スコアと単語の音響スコアを任意の規定の割合で加算して照合結果とするものである。

【0026】

【発明の実施の形態】以下図面を参照しながら、この発明の実施の形態を説明する。

【0027】実施の形態1. 図9との対応する部分に同一符号を付けた図1に、この発明による実施の形態1の

$$S(w) = -1.0 \times \log(P(w))$$

【0032】式(1)においてwは単語、P(w)は単語wの使用頻度を確率で表したものの、S(w)は単語wの使用頻度スコアである。S(w)は使用頻度が高い単語には小さな値、使用頻度の大きな単語には大きな値となる。しかし、あまりに使用頻度が小さすぎると使用頻度スコアが非常に大きな値となり、どんなに丁寧に発声しても結果出力部4において上位の認識結果として出力されなくなるため、使用頻度スコアの下限値を設けることにより、非常に使用頻度の少ない単語でも、音響スコアが小さければ認識結果として出力することが可能となる構成としても良い。

音声認識装置を示す。図9について上述した従来の音声認識装置と同様に、音響分析部1は入力される音声信号S1を一定時間毎に音響分析し、特徴パラメータベクトルS2と音声信号のパワーS3に変換し出力する。音声区間検出部2は音響分析部1から受け取る音声信号のパワーS3の変化により音声信号S1の音声区間を検出し、音声区間の検出状況の変化により音声区間検出信号S4を出力する。

【0028】照合部3は音声区間検出信号S4の指示に従い音響分析部1から受け取る特徴パラメータベクトルの時系列のうち音声区間のものと、単語辞書記憶部10から読み出される順番でスコア付単語辞書S12との照合を行い、入力された音声信号S1とスコア付単語辞書S12がどの程度近いかを示す音響スコアと使用頻度スコアをある規定の割合で加え距離値S5として順次出力する。ここでこの実施の形態1の場合、頻度付単語辞書記憶部7は認識対象単語の読みを表すラベルと使用頻度を表す情報を収め、使用頻度スコア計算部11は頻度付単語辞書記憶部7の使用頻度を表す情報に従い頻度付き単語辞書S8に使用頻度スコアを付加し、使用頻度の高い順番にスコア付単語辞書S11として出力する。単語辞書記憶部10は使用頻度の高い順番にスコア付単語辞書S11を記憶する。

【0029】結果出力部4は音声区間検出信号S4又は外部から入力される出力要求信号S9が入力された時、既に受け取った距離値S5のうちまだ出力していないものをソーティングして距離値S5の小さな1つ又は複数の単語を認識結果S10として出力する。

【0030】このような構成の音声認識装置の動作について説明する。認識に先立って、使用頻度スコア計算部11では頻度付単語辞書記憶部7の内容を読みだし、使用頻度からスコア付単語辞書S11を求め、使用頻度の高い順番に単語辞書記憶部10に記憶する。この使用頻度スコアの与えかたとしては、例えば次式のような演算式により求める方法がある。

【0031】

【数1】

$$\dots\dots (1)$$

【0033】この音声認識装置の認識時の動作について説明する。音響分析部1、音声区間検出部2、結果出力部4の動作は、図9～図11について説明した従来の音声認識装置と同様のためここでは説明を省略する。ここではこの実施の形態1の特徴である照合部3の動作について説明する。照合部3では従来の音声認識装置と同様に単語辞書記憶部10のスコア付単語辞書S12を順番に読みだしパターン照合を行うが、次式に示すように音響スコアD(w)に対し使用頻度スコアS(w)を重みRで加える。

【0034】

【数 2】

$$D'(w) = D(w) + R \times S(w)$$

..... (2)

【0035】これにより、使用頻度スコアの低い単語は認識しやすくし、スコアの高い単語は認識しづらくする。すなわち、使用頻度の高い単語は認識しやすくし、使用頻度の低い単語は認識しづらくする効果を与える。この実施の形態1によれば、図11について上述した認識実験と同じ条件による認識実験を行った結果、57,711単語を認識対象語彙とした時の誤り率を63.4%から32.1%に改善できた。

【0036】実施の形態2. 上述の実施の形態1では、使用頻度スコア計算部11の機能として使用頻度の高い順番に並べて単語辞書記憶部10に収めるとして説明をしたが、H/Wが十分に速く全単語候補に対するパターン照合が高速に処理可能な場合や、加えて、「孤立単語音声認識における全探索法・ビームサーチ法・A*探索法の比較」(平成8年度春季日本音響学会講演論文集、2-5-10、伊田正樹、中川聖一著)に記載されているビームサーチ法や枝刈り法に代表される演算量削減策を講じることにより、図10におけるフレームT5までにパターン照合で全候補の照合結果を得られる場合には、従来の音声認識装置のように分割してパターン照合を行う必要はなく、加えて使用頻度スコア計算部11の機能として使用頻度の高い順番に並べて単語辞書記憶部7に収める必要はない。

【0037】このような高速にパターン照合の処理が可能な照合部3を有する音声認識装置を実施の形態2に示す。この音声認識装置の構成は実施の形態1と同様なので、ここでは説明を省略する。このような構成の音声認識装置の動作について説明する。認識に先立って、使用頻度スコア計算部11では頻度付単語辞書記憶部7の内容を読み出し、使用頻度からスコア付単語辞書S11を求め、単語辞書記憶部10に記憶する。単語辞書記憶部10には使用頻度の高い順番に並べることは必要なくランダムに並べて良い。この使用頻度スコアの与え方としては、上述した実施の形態1と同様である。

【0038】音響分析部1、音声区間検出部2の動作は、図9～図11について上述した従来の音声認識装置と同様であり、ここでは説明を省略する。図2はこの実施の形態2に基づく音声認識装置の動作を説明するタイミングチャートである。以下この図2を用いて、照合部3及び結果出力部4の動作について説明する。フレームT5より前の処理は従来の音声認識装置と同様である。この実施の形態2による照合部3では十分に処理能力が高いため、フレームT5以前にパターン照合の処理を終了している。そのためフレームT5において、結果出力部4は音声区間検出部2の音声区間検出信号S4としての端末確定信号により、照合部3から転送された距離値S5をソーティングして複合スコアの小さい1つ又は複数の単語の認識結果S10として出力する。さらに外部か

らの出力要求信号S9があった場合には、先に出力した認識結果S10を除いて、さらに距離値S5の小さい1つ又は複数の単語を認識結果S10として出力する。

【0039】実施の形態3. 上述した実施の形態1、実施の形態2では単語の端末候補が定まった後、一単語ずつパターン照合を行う方式の音声認識装置について述べてきたが、フレーム同期型パターン照合を行う照合部3を用いても、同様の効果を実現できる。フレーム同期型パターン照合は全単語辞書に対するパターン照合を同時に進めて行く方法である、一単語ずつパターン照合を行う方法に比べ、ワークメモリ量は大きく増加するという欠点はあるが、音声入力と平行してパターン照合を行えるため、パターン照合を効率的に行えるという特徴を持つ。フレーム同期パターンパターン照合は、例えば「フレーム同期化、ビームサーチ、ベクトル量子化の統合によるDPマッチングの高速化」(電子通信学会論文誌D、Vol. J71-D、No. 9、pp1650-1659、迫江博昭、藤井浩美、吉田和永、亘理誠夫共著)等に記述されている方法を用いる。

【0040】このような音声認識装置の構成は実施の形態1の構成と同じであり、ここでは説明を省略する。この実施の形態3としての音声認識装置の動作について説明する。音響分析部1及び音声区間検出部2の動作は実施の形態2と同じなので、ここでは説明を省略する。照合部3及び結果出力部4の動作について、図3を用いて説明する。まず照合部3の動作について、照合部3は音声区間検出部2からの音声区間検出信号S4の1つである始端信号によりパターン照合処理を開始し、音声区間検出信号S4の1つである端末確定信号により動作を終了する。

【0041】距離値S5は毎フレームにおいて照合部3から出力される。結果出力部4は音声区間検出信号S4の1つである端末候補信号により、端末候補のフレームの距離値S5をソーティングし、音声区間検出信号S4の1つである端末確定信号により距離値11の小さな1つ又は複数の単語を認識結果S10として出力する。図3には、フレームT2、T4の2つの端末候補信号があるが、フレームT5において出力するのは、フレームT4において得られた認識結果S10である。このようにフレーム同期型のパターン照合を行うことにより、従来の音声認識装置では演算を行っていなかったフレームT1からフレームT2及びフレームT3からフレームT4においても照合部3の処理を行うことができ、効率的な音声認識装置を実現できる。

【0042】実施の形態4. 上述の説明では使用頻度が既知の単語について述べたが、音声認識装置をある程度運用すれば、使用頻度が得られる場合があるが、運用初期の段階では使用頻度を得ることは難しい場合が多い。

しかし、例えば自治体にある住民台帳や企業の持つ顧客データベースや社員データベースには、住所、姓名、電話番号、性別、年齢等が記録されている。そのため例えば、住民に対する情報サービスシステムなどでは、住民台帳の内容から単語の使用頻度が推定できる。つまり人口の割合の多い住所、姓名、電話番号等の単語は使用頻度は高いと推定する。企業の持つ顧客データベースや社員データベースに対しても同様の推定が可能である。この実施の形態 4 では、姓をひとつの例として使用頻度を推定する音声認識装置について説明する。

【0043】この実施の形態 4 による音声認識装置の構成を図 1 との対応部分に同一符号を付けて図 4 に示す。図 4 において、音響分析部 1、音声区間検出部 2、照合部 3、結果出力部 4、頻度付単語辞書記憶部 7、使用頻度スコア計算部 11、単語辞書記憶部 10 は実施の形態 3 と同様なので、ここでは説明を省略する。図におい

$$P(w) = N(w) / (ALLN)$$

【0046】式 (3) において、ALLN はそのデータベース 12 に含まれる全人口である。単語の読みはデータベース 12 に含まれる読みを用いる。その後のスコア付単語辞書 S 11 の作成方法や音響分析部 1、音声区間検出部 2、照合部 3、結果出力部 4 の動作は実施の形態 3 に等しいのでここでは説明を省略する。

【0047】実施の形態 5、上述した実施の形態 4 では頻度付単語辞書記憶部 7 の単語の読みをデータベース 12 に含まれるふり仮名を用いる例を述べたが、一般的にデータベース 12 に含まれるふり仮名は書く時のかな文字がふられており、音声認識装置に入力される発声とは一致しない場合がある。例えばデータベース 12 における「佐藤」のかな文字表記は「さとう」であるが、約 80 % の人はこれを「さとー」と長音で発声する。残り約 15 % の人は文字通り「さとう」と発声する。「さ、と、う」と区切って発声する人も存在する。これらは全て「佐藤」と音声認識すべきなので、自動的にこれらの単語を加え、使用頻度を推定することにより、認識率を向上させることが可能となる。

【0048】この実施の形態 5 の音声認識装置の構成は実施の形態 4 の図 4 に等しいので、ここでは説明を省略する。ただし、図 4 の使用頻度推定部 13 の動作は実施の形態 4 に示したものと異なり、単語の変形規則を用いて同じふり仮名に対する異なる読み方の単語を追加する機能を持つ。以下この実施の形態 5 における使用頻度推定部 13 の動作を説明する。図 5 はこの発明における使用頻度推定部 13 の動作を示す流れ図である。図において処理は「START」から始まり「END」で終

る。まず図中ステップ S T 1 においてローマ字表記で「OU」が含まれる単語に対しては、ステップ S T 2 において「OU」を「OO」に変えた単語を追加する。使用頻度はもとの単語の値に対し 0.8 の倍率を乗ずる。もとの「OU」を含む単語の使用頻度は 0.2 の倍率を乗じ

て、データベース 12 は住民の姓が含まれているデータベースであり、この中には姓に対してかな文字でふり仮名がふられているものとする。また使用頻度推定部 13 はデータベース 12 から名の頻度情報と読みを生成するものである。さらに S 13 は姓情報、S 14 は頻度付き単語辞書である。

【0044】頻度付単語辞書記憶部 7 の推定方法について説明する。まずデータベース 12 を検索して、同じふり仮名を持つ姓をひとつの単語 w として、単語 w に対する人口 N (w) を調べる。異なる漢字であってもふり仮名が同じであれば同一の単語 w として計数する。そしてそのような姓を持つ人が多ければ、その姓の使用頻度も高いと推定し、使用頻度 P (w) を次式で求める。

【0045】

【数 3】

..... (3)

る。

【0049】次にステップ S T 3 において、ローマ字表記で「EI」が含まれる単語に対しては、ステップ S T 4 において「EI」を「EE」に変えた単語を追加する。使用頻度は元の単語の確率に対し 0.7 の倍率を乗ずる。元の「EI」を含む単語の使用頻度は 0.3 の倍率を乗じる。またステップ S T 5 において、促音を含む単語であったなら人により促音を「つ」と発声することがあるため、ステップ S T 6 において促音を「つ」に変えた単語を追加する。使用頻度は元の単語の値に対し、0.05 の倍率を乗じる。元の促音を含む単語の使用頻度は、0.95 の倍率を乗じる。

【0050】次にステップ S T 7 において、全単語に対し長音化した単語と切断化した単語を追加する。元の単語の使用頻度に対し長音化した単語は 0.1 の倍率を乗じ、切断化した単語には 0.05 の倍率を乗じ使用頻度とする。元の単語の使用頻度は 0.85 の倍率を乗じて変更する。ただし、母音及び撥音の長音化においては最後の音節は長音化しない場合もあるため、このような変形規則を用いても良い。

【0051】このような構成の音声認識装置による具体的な処理結果を示す。図 6 はあるデータベース 12 を用いたときの実施の形態 4 に示された使用頻度推定部 13 で推定される単語と使用頻度であるとする。これに対し、この実施の形態 5 における使用頻度推定部 13 では図 7 に示す 20 個の単語が推定される。図 7 において、ハイフン (-) は母音及び撥音が長音化されていることを示し、点 (.) は音節が切断されていることを示す。母音及び撥音の長音化では、最期の音節は長音化しないという変形規則を用いている。

【0052】各単語の使用頻度は、図 5 の流れに従い規定の倍率を掛けられている。例えば「あべ」は図 5 のステップ S T 7 の規則を適用されて母音が長音化された

「あーべ」と音節毎に切断された「あ・べ」が追加される。使用頻度は元の使用頻度 0.04598 に対し「あべ」が 0.85 倍、「あーべ」が 0.10 倍、「あ・べ」が 0.05 倍されている。ただし、「にった」に関しては音節毎に切断された単語と、もとの単語が同じとなるため、「にった」の使用頻度が 0.90 倍されている。

【0053】このようにこの実施の形態 5 によれば、データベース 12 のふり仮名から、様々な発声の変形とその使用頻度を推定するため、良好な認識性能を示す音声認識装置を実現できる。なお上述の倍率の値は任意の調査結果から経験的に求めたものであるが、これらはデータベースに応じて変更しても良い。

【0054】実施の形態 6. データベースの内容において人口に偏りがある場合がある。例えば名前では男性、女性で人口が異なる。そのため、音声信号が男性か女性かの情報を追加すればさらに認識性能を高めることができる。この実施の形態 6 の音声認識装置の構成を図 8 に示す。図において、音響分析部 1、音声区間検出部 2、照合部 3、結果出力部 4、頻度付単語辞書記憶部 7、使用頻度スコア計算部 11、単語辞書記憶部 10、データ

ベース 12、使用頻度推定部 13 の構成は、上述した実施の形態 5 と同様である。

【0055】図 8 において話者識別部 14 は音声区間検出部 2 からの音声区間検出信号 S4 に従い、音響分析部 1 からの特徴パラメータベクトル S2 を比較し話者識別

$$S2(i) = \frac{\sum_{t=T1}^{T4} \min(dis(M(i, m), L(t)))_{m=1, M}}{T4 - T1 + 1}$$

..... (4)

【0059】式 (4) において、次式

【0060】

$$\min(X(m))_{m=1, M}$$

..... (5)

【0061】は要素 X (m) の m = 1, M に関する最小値を意味する。また、

$$dis(M(i, m), L(t))$$

【0063】は M (i, m) と L (T) の距離値を意味する。式 (4) の演算はフレーム T4 においてまとめて行うことも可能であるし、フレーム T1 からフレーム同

期的に行うことも可能である。フレーム T1、T4 は音声区間検出信号 S4 として通知される。このようにして得られた話者識別スコア S15 は照合部 3 において R2

$$D'(w) = \min(D(w) + R \times S1(i, w) + R2 \times S2(i))_{i=1, 2}$$

..... (7)

【0065】式 (7) において D (w)、R は式 (2) で用いたものと同じであり、S1 (i, w) は性別 i の単語 w に対する使用頻度スコアである。

【0066】上述の説明では話者識別部 14 の標準パ

を行うものである。ここでは話者識別の対象を、男性、女性として、性別により姓名のうち名前を男性、女性で分類して記憶しておくことを一例として説明する。

【0056】まず、認識に先立ち使用頻度推定部 13 では、データベース 12 の同じ名前に対し男性、女性に分けて別の単語として頻度付単語辞書 S14 を作成する。以下、使用頻度スコア計算部 11 でも、別々の単語としてスコアを計算し、単語辞書記憶部 10 に格納する。これにより、頻度付単語辞書記憶部 7、単語辞書記憶部 10 の記憶量は 2 倍になる。また話者識別部 14 には話者識別のための標準パターンが入れられる。話者識別の方法としては多くのものが提案されているが、ここではベクトル量子化を用いる方法を一例として説明する。

【0057】話者識別部 14 には男性用を 1、女性用を 2 としてそれぞれ M 個の標準パターンを用意する。この標準パターンは男性、女性それぞれの音声信号から LBG (Linde Buzo Gray) アルゴリズム等を用いて学習される。性別 i の m 番目の標準パターンを M (i, m)、フレーム t の特徴パラメータベクトル 9 を L (t) とすると、話者識別スコア 27 である S2 (i) は、次式の演算式で求められる。

【0058】

【数 4】

【数 5】

【0062】

【数 6】

..... (6)

の割合で音響スコアと使用頻度スコアに加えられ、男性用のものと女性用のものと小さなほうが最終的な照合結果となる。

【0064】

【数 7】

ーンを単語辞書記憶部 10 の標準パターンとは別のものとして説明したが、男性用、女性用の単語辞書記憶部 10 の標準パターンを持つマルチテンプレートの音声認識装置では、これを話者識別用に流用することも可能であ

り、このような構成でも上述と同様の効果を実現できる。また上述の説明では性別により話者識別を行う例を述べたが、年齢や日本人名か英語名かの言語などによりデータベース 1 2 を分類して話者識別を行うことも可能であり、同様に効果を実現できる。

【 0 0 6 7 】

【発明の効果】以上のようにこの発明によれば、単語の使用頻度から計算した使用頻度スコアを音響スコアに規定の割合で加えて距離値を得るため、頻度の多い単語の認識性能を高めることができ、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【 0 0 6 8 】また次の発明によれば、非常に使用頻度の低い単語のスコアの下限値を設けたため、極めて使用頻度の低い単語でも音響スコアが良好であれば、上位の認識結果とでき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【 0 0 6 9 】また次の発明によれば、既存のデータベースから使用頻度を推定できるため、使用頻度が明確でない単語に対しても、使用頻度スコアを設定でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【 0 0 7 0 】また次の発明によれば、既存のデータベースのふり仮名にローマ字表記で「OU」を含む単語がある時、これを「OO」と変更した単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【 0 0 7 1 】また次の発明によれば、既存のデータベースのふり仮名にローマ字表記で「EI」を含む単語がある時、これを「EE」と変更した単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【 0 0 7 2 】また次の発明によれば、既存のデータベースのふり仮名に対し母音を長音化した単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【 0 0 7 3 】また次の発明によれば、既存のデータベースのふり仮名に対し音節毎に区切られた単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【 0 0 7 4 】また次の発明によれば、既存のデータベー

スのふり仮名に促音を含む単語がある時、これを「つ」と変更した単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【 0 0 7 5 】また次の発明によれば、規定の基準においてデータベースの内容を分類して使用頻度を推定し、認識時には話者照合を行い、話者照合スコアを使用頻度スコアと音響スコアにある割合で加えるため、良好な認識性能を得ることができ、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【図面の簡単な説明】

【図 1】 この発明による音声認識装置の実施の形態 1 の構成を示すブロック図である。

【図 2】 この発明による音声認識装置の実施の形態 2 の動作の説明に供するタイミングチャートである。

【図 3】 この発明による音声認識装置の実施の形態 3 の動作の説明に供するタイミングチャートである。

【図 4】 この発明による音声認識装置の実施の形態 4 の構成を示すブロック図である。

【図 5】 この発明による音声認識装置の実施の形態 5 の使用頻度推定部の動作を示すフローチャートである。

【図 6】 この発明による音声認識装置の実施の形態 4 の使用頻度推定部の処理結果の説明に供する図表である。

【図 7】 この発明による音声認識装置の実施の形態 5 の使用頻度推定部の処理結果の説明に供する図表である。

【図 8】 この発明による音声認識装置の実施の形態 6 の構成を示すブロック図である。

【図 9】 従来の音声認識装置の構成を示すブロック図である。

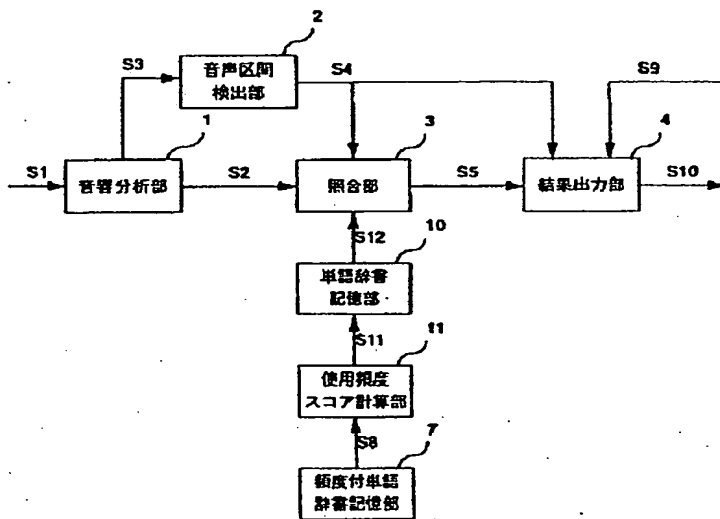
【図 1 0】 図 9 の音声認識装置における音声区間検出部の動作の説明に供するタイミングチャートである。

【図 1 1】 従来の音声認識装置による認識性能の説明に供する特性曲線図である。

【符号の説明】

- 1 音響分析部
- 2 音声区間検出部
- 3 照合部
- 4 結果出力部
- 5 単語辞書記憶部
- 6 単語辞書ソーティング部
- 7 頻度付単語辞書記憶部
- 1 0 単語辞書記憶部
- 1 1 使用頻度スコア計算部
- 1 2 データベース
- 1 3 使用頻度推定部
- 1 4 話者識別部

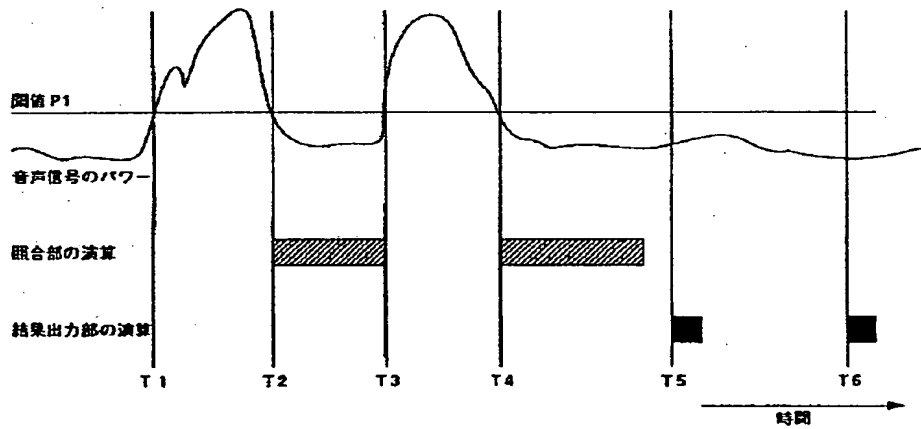
【図 1】



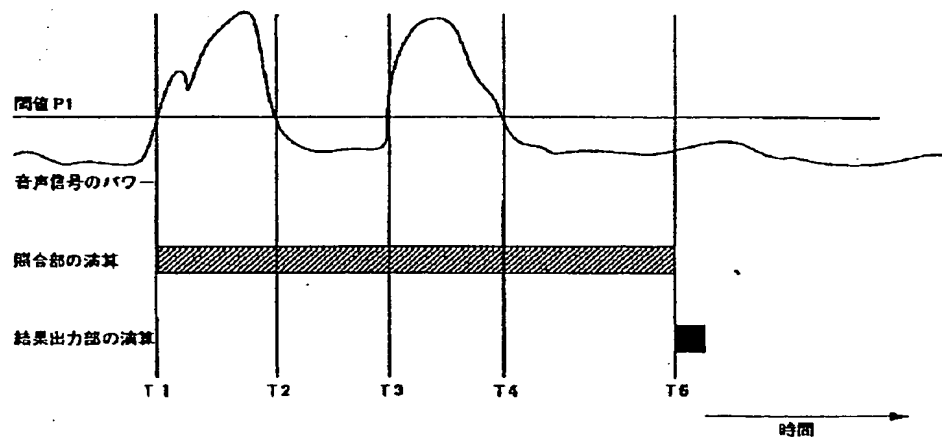
【図 6】

単語	使用頻度
あべ	0.04598
さとう	0.04054
にった	0.00066
せいの	0.00076

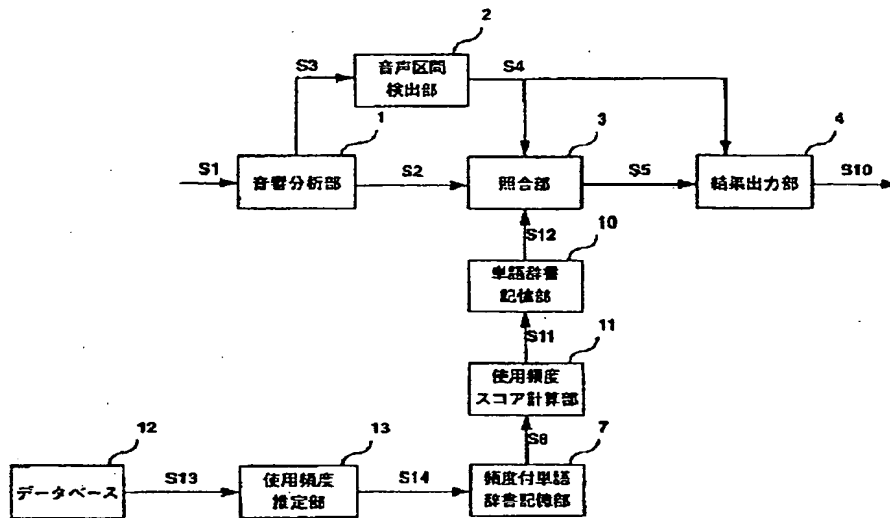
【図 2】



【図 3】



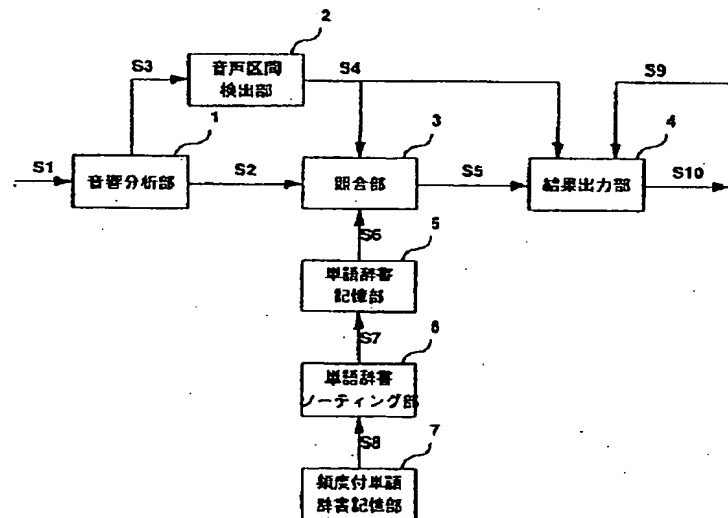
【図 4】



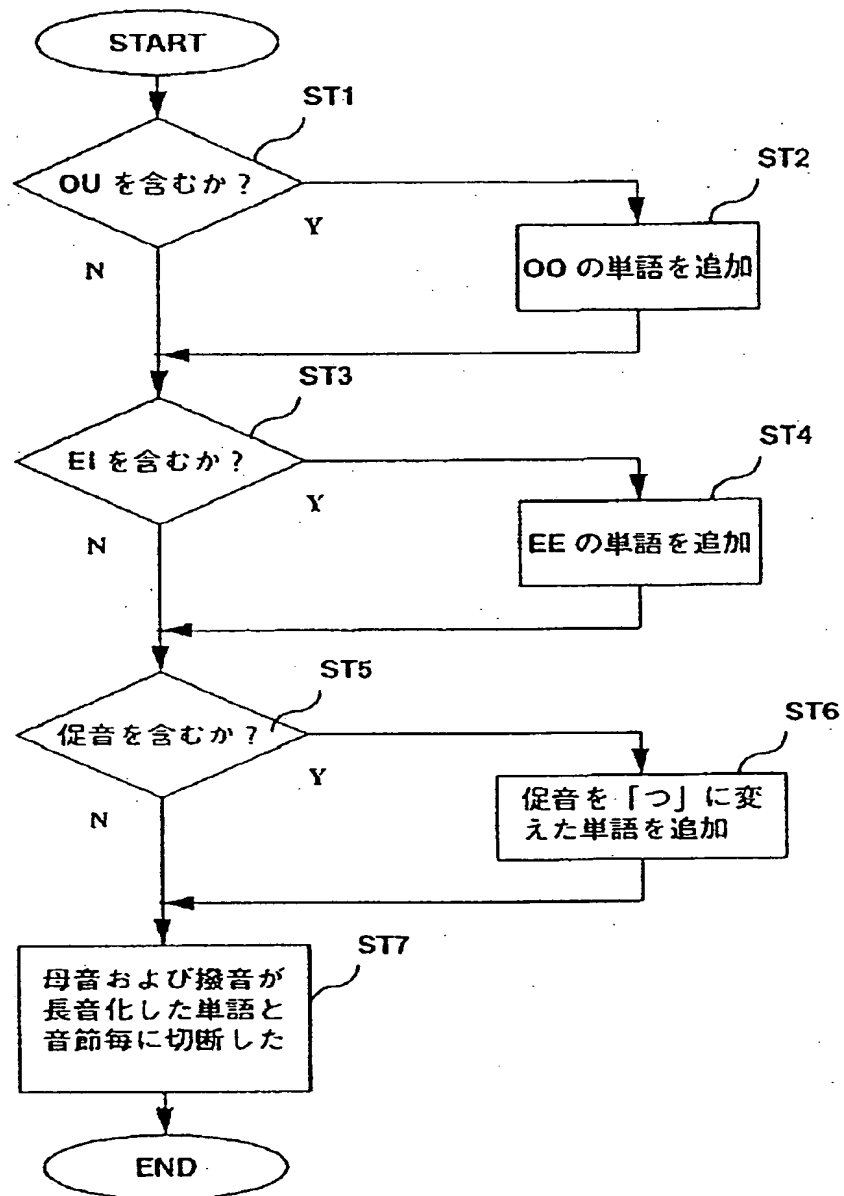
【図 7】

単語	使用頻度
あべ	0.04598×0.85
あーべ	0.04598×0.10
あ・べ	0.04598×0.05
さとう	$0.04054 \times 0.2 \times 0.085$
さとお	$0.04054 \times 0.8 \times 0.085$
さーとう	$0.04054 \times 0.2 \times 0.010$
さーとお	$0.04054 \times 0.8 \times 0.010$
さ・と・う	$0.04054 \times 0.2 \times 0.005$
さ・と・お	$0.04054 \times 0.8 \times 0.005$
にった	$0.00086 \times 0.95 \times (0.85 + 0.05)$
につた	$0.00086 \times 0.05 \times 0.85$
にーった	$0.00086 \times 0.95 \times 0.10$
にーつた	$0.00086 \times 0.05 \times 0.10$
に・つ・た	$0.00086 \times 0.05 \times 0.05$
せいの	$0.00076 \times 0.6 \times 0.85$
せえの	$0.00076 \times 0.4 \times 0.85$
せーいの	$0.00076 \times 0.6 \times 0.10$
せーえの	$0.00076 \times 0.4 \times 0.10$
せ・い・の	$0.00076 \times 0.6 \times 0.05$
せ・え・の	$0.00076 \times 0.4 \times 0.05$

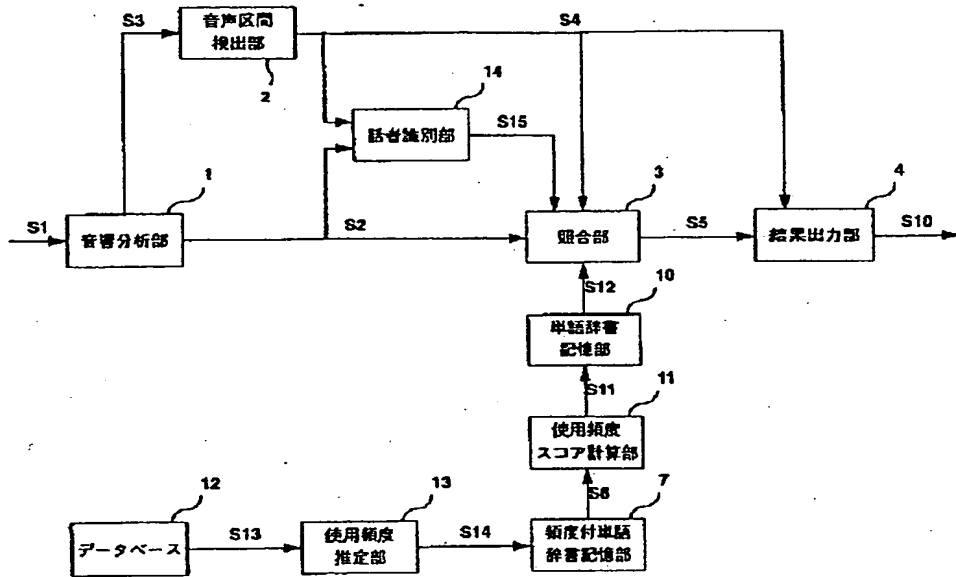
【図 9】



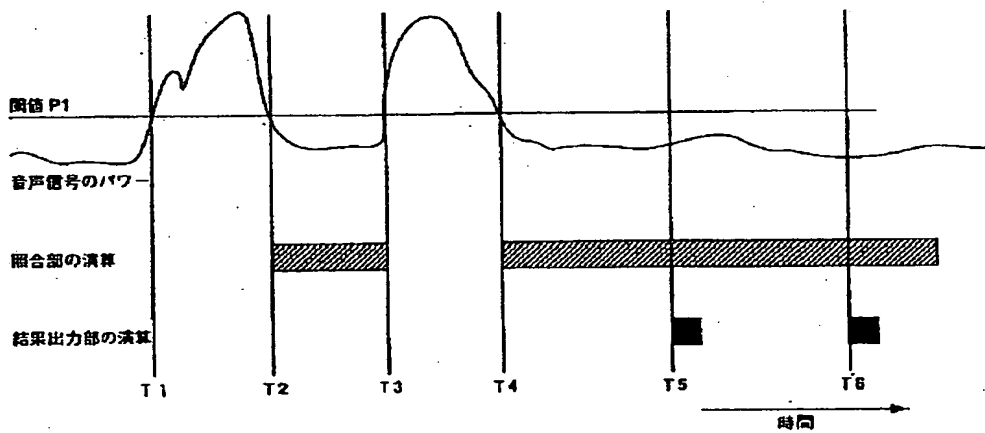
【図 5】



【図 8】



【図 10】



【図 1 1】

